No. K-2/68

Technical Memorandum

A COMPUTER PROGRAM FOR THE KOLMOGOROV GOODNESS OF FIT TEST FOR NORMALITY

Carl B. Bates
Jacqueline R. Orsulak

Computation and Analysis Laboratory

U. S. Naval Weapons Laboratory
Dahlgren, Virginia

54

U. S. NAVAL WEAPONS LABORATORY

TECHNICAL MEMORANDUM

January 1968

No. K-2/68

A COMPUTER PROGRAM FOR THE KOLMOGOROV GOODNESS OF FIT TEST FOR NORMALITY

Carl E. Bates
and
Jacqueline R. Orsulak

Computation and Analysis Laboratory

Approved by:

H. S. OVERMAN, Acting
Director, Computation and
Analysis Laboratory

While the contents of this memorandum are considered to be correct, they are subject to modification upon further study.

Distribution of this document is unlimited.

TABLE OF CONTENTS

NWL Technical Memorandum No. K-2/68

Initial Distribution:

Commander
Naval Air Systems Command (Code 604)
Washington, D. C. 20360 (2)

Commander
Naval Ordnance Systems Command (Code 9132)
Washington, D. C. 20360 (2)

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314 (20)

Local:

K	(1)
K-1	(1)
KE	(1)
KG	(1)
KP	(1)
KPO	(2)
KR	(1)
KRX	(1)
KRM	(25)
KRS	(1)
KRT	(1)
KRU	(1)
KW	(1)
KYD-1	(1)
T	(1)
TDM	(1)
TE	(1)
WDA	(1)
WHY	(1)
WW-2	(1)
WWI	(1)
MAL	(6)
MAM	(1)

ABSTRACT

Following a brief discussion on the background, applicability, and limitations of the Kolmogorov Test, a computer program of the Kolmogorov Test for normality is described. The program is applicable for testing the null hypothesis that a random sample is from a parent normal population whose mean and variance are equal to those of the sample distribution. The program determines the minimum and maximum sample values, computes sample estimates of the mean and variance of the hypothesized normal distribution, and computes the Kolmogorov statistic. Five optional significance levels (0.20, 0.15, 0.10, 0.05, and 0.01) are available, and sample size limitations are $4 \leq n \leq 2500$. An optional feature provides for CRT plot output applicable for both hypothesis testing and interval estimation.

The program is coded in FORTRAN IV for the IBM 7030 (STRETCH) computer.

FOREWORD

The formulation of the computer program for the Kolmogorov Test was performed in the Mathematical Statistics Branch of the Operations Research Division. The coding of the program was performed in the Operations Sciences Branch of the Computer Programming Division. Dr. H. W. Lilliefors of the George Washington University supplied the critical values which are tabulated within the program. The project was supported under Foundational Research Project No. 29Y, "Computer Programs for Statistical Analyses."

The date of completion was 15 December 1967.

I. INTRODUCTION

The applied statistician is often confronted with the problem of investigating the distribution of sampled data. Because of the desirable characteristics of the normal distribution and because of the frequency of occurrence of these characteristics in physical phenomena, a logical approach to the problem is an examination of the sample distribution to determine if it can be approximated by a normal distribution. This examination is often accomplished by hypothesis testing, i.e., testing the null hypothesis that the randomly observed sample data is from a parent normal population.

The Chi-Square Test originally proposed by Pearson in 1900 appears to be the most popular of the Goodness of Fit Tests. A common criticism, however, of the Chi-Square Test is the loss of power because of the required grouping of the sample data, e.g., see Massey (1951), Birnbaum (1952), or Cochran (1954). In addition, to employ the Chi-Square Test to test for normality, a minimum of four intervals (groups) is needed when the population is not completely specified under the null hypothesis and the two population parameters (μ and σ^2) are estimated by sample statistics. Consequently, the Chi-Square Test is not applicable to very small samples.

An alternative distribution-free test proposed by Kolmogorov in 1933 does not have the objectionable requirements mentioned above for the Chi-Square Test. That is, the Kolmogorov Test requires no grouping of the data and the test is applicable to very small samples. Because the Kolmogorov Test treats individual observations rather than grouped data, sample information may be better utilized by the Kolmogorov Test than by the Chi-Square Test, see Birnbaum (1952). Studies by Massey (1951) and

Kac, Kiefer, and Wolfowitz (1955) indicate that the asymptotic power of the Kolmogorov Test is greater than that of the Chi-Square Test. A recent paper by Slakter (1967), however, shows that the Kolmogorov Test is not uniformly more powerful than the Chi-Square Test. Further studies are necessary, in the opinion of the author of this report, to justify a preference of either test on the basis of power. On the other hand, the definite advantage of the Kolmogorov Test with respect to small samples is sufficient justification for the applied statistician to give serious consideration to the Kolmogorov Test.

II. THE KOLMOGOROV TEST

A. Background

The Kolmogorov Test for normality, like the Chi-Square Test for normality, is a test for assessing the agreement of a sample distribution with a parent normal distribution. While the Chi-Square Test is concerned with the agreement of a sample distribution with a theoretical distribution, the Kolmogorov Test is concerned with the agreement of a sample cumulative distribution with a theoretical cumulative distribution.

Consider a random variable X with a continuous cumulative distribution function (cdf), F(x), where

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(x)\,dx. \tag{1}$$

Let $x_1, x_2, \cdots, x_i, \cdots, x_n$ be a random sample of n observations (or measurements) of X. The empirical cdf corresponding to the random sample of size n is $S_n(x)$, where

$$S_n(x) = \frac{\text{number of } x_i \le x}{n}. \tag{2}$$

The Kolmogorov statistic, $D(n)$, is based on the maximum absolute deviation between $F(x)$ and $S_n(x)$. That is,

$$D(n) = \text{Max} \left| F(x) - S_n(x) \right| . \tag{3}$$

$D(n)$ then is the test statistic to test the null hypothesis of no difference between $S_n(x)$ and $F(x)$.

Until recently the Kolmogorov Test was applicable for only completely specified $F(x)$ under the null hypothesis. That is, the standard tables, such as Massey (1951), Birnbaum (1952), and Miller (1956), of critical values for testing $D(n)$ are not applicable if population parameters of the theoretical distribution are estimated by sample statistics. Lilliefors (1967) provided a tabulation of critical values applicable when the theoretical distribution is normal and sample statistics are used to estimate the two population parameters. The critical values, which were determined by Monte Carlo calculations, are for five significance levels (0.20, 0.15, 0.10, 0.05, and 0.01) for $n \geq 4$. Consequently, the Kolmogorov Test can now be used to examine a sample distribution to determine if it can be approximated by a normal distribution whose mean and variance are equal to those of the sample distribution.

B. The Kolmogorov Test for Normality

To distinguish the incompletely specified from the completely specified theoretical distribution under the null hypothesis, a circumflex is used denoting the use of sample statistics for the two population parameters of the normal distribution. Equation (1) then becomes

$$\hat{F}(x) = P[X \le x] = \int_{-\infty}^{x} \hat{f}(x)\,dx, \tag{4}$$

where

$$\hat{f}(x) = (1/\sqrt{2\pi}s)\ \exp[-(x-\bar{x})^2/2s^2], \tag{5}$$

and where the familiar statistics, $\bar{x}$ and s^2, are

$$\bar{x} = \sum_{i=1}^{n} x_i/n \qquad \text{and} \qquad s^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1). \tag{6}$$

Substituting $\hat{F}(x)$ for $F(x)$ in equation (3), the Kolmogorov statistic becomes

$$\hat{D}(n) = \mathrm{Max}\,\left|\hat{F}(x) - S_n(x)\right|. \tag{7}$$

The null hypothesis that a random sample of n observations is from a parent normal population is tested by comparing $\hat{D}(n)$ with the critical value, $D_\alpha(n)$, of TABLE I. If $\hat{D}(n) \ge D_\alpha(n)$, reject the null hypothesis at the α-level of significance; otherwise, do not reject the null hypothesis.

TABLE I

$D_\alpha(n)$ CRITICAL VALUES*

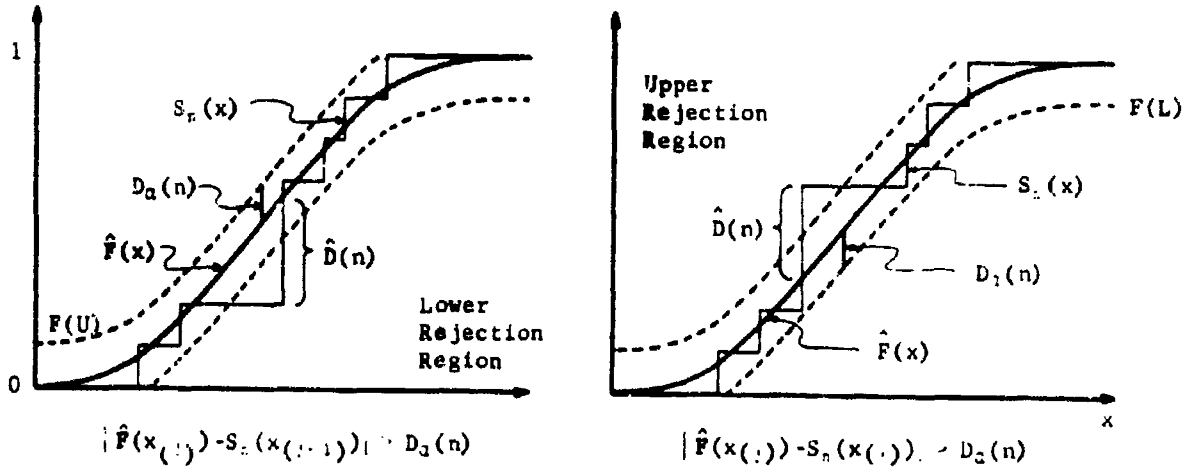
Sample Size n	α - Level of Significance 0.20	0.15	0.10	0.05	0.01
4	.300	.319	.352	.381	.417
5	.285	.299	.315	.337	.405
6	.265	.277	.294	.319	.364
7	.247	.258	.276	.300	.348
8	.233	.244	.261	.285	.331
9	.223	.233	.249	.271	.311
10	.215	.224	.239	.258	.294
11	.206	.217	.230	.249	.284
12	.199	.212	.223	.242	.275
13	.190	.202	.214	.234	.268
14	.183	.194	.207	.227	.261
15	.177	.187	.201	.220	.257
16	.173	.182	.195	.213	.250
17	.169	.177	.189	.206	.245
18	.166	.173	.184	.200	.239
19	.163	.169	.179	.195	.235
20	.160	.166	.174	.190	.231
25	.142	.147	.158	.173	.200
30	.131	.136	.144	.161	.187
Over 30	$\frac{.736}{\sqrt{n}}$	$\frac{.768}{\sqrt{n}}$	$\frac{.805}{\sqrt{n}}$	$\frac{.886}{\sqrt{n}}$	$\frac{1.031}{\sqrt{n}}$

*From Lilliefors (1967), On The Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, Journal of the American Statistical Association, Vol. 62, pp. 399-402 and personal communication of October 16, 1967

The Kolmogorov Test is a two-sided test in that departures of $S_n(x)$ in either direction from $\hat{F}(x)$ increase $\hat{D}(n)$. The rejection region of $\hat{D}(n)$ is the region __outside__ of the band $\hat{F}(x) \pm D_\alpha(n)$, i.e.,

$$
\begin{aligned}
F(U) &= \hat{F}(x) + D_\alpha(n) \\
F(L) &= \hat{F}(x) - D_\alpha(n).
\end{aligned}
\tag{8}
$$

That is, the null hypothesis is rejected at the α-level of significance if $S_n(x)$ passes outside this band. Because $S_n(x)$ is a step-function, $S_n(x)$ is constant for $x_{(j-1)} \leq X < x_{(j)}$, where $x_{(j)}$ is the j^{th} order statistic. Therefore, if $S_n(x)$ passes into the lower rejection region, $\hat{D}(n) = |\hat{F}(x_{(j)}) - S_n(x_{(j-1)})| \geq D_\alpha(n)$, see Figure 1(a); if $S_n(x)$ passes into the upper rejection region, $\hat{D}(n) = |\hat{F}(x_{(j)}) - S_n(x_{(j)})| \geq D_\alpha(n)$, see Figure 1(b).



(a) (b)

Figure 1

In addition to the utility of $D_\alpha(n)$ for hypothesis testing, $D_\alpha(n)$ may also be used for interval estimation. The $100(1-\alpha)\%$ confidence limits for the "true" cumulative distribution function, $F(x)$, are

$$S(L) < F(x) < S(U), \tag{9}$$

where

$$\begin{aligned} S(U) &= S_n(x) + D_\alpha(n) \\ S(L) &= S_n(x) - D_\alpha(n). \end{aligned} \tag{10}$$

III. COMPUTATIONAL PROCEDURE

Before performing the computations described in the previous section, the n sample values are transformed by one of the following thirteen available transformations.

Transformation Number	Transformation
1	$x \leftarrow x$
2	$x \leftarrow \ln x$
3	$x \leftarrow \ln(\ln(x))$
4	$x \leftarrow \ln(A+x)$
5	$x \leftarrow \ln(B+\ln(C+x))$
6	$x \leftarrow \sqrt{x}$
7	$x \leftarrow 1/x$
8	$x \leftarrow 1/(D+x)$
9	$x \leftarrow \sin^{-1} x$
10	$x \leftarrow 2 \sin^{-1}\sqrt{x}$
11	$x \leftarrow x/E$
12	$x \leftarrow \sin x$
13	$x \leftarrow \cos x$

The transformations are identified on card type 3, and the constants A,B,C,D, and E are input on card type 5 (see Section IV).

After ordering the n transformed sample values, the subset

$$x_{(1)} < x_{(2)} < \cdots < x_{(j)} < \cdots < x_{(k)} \; ; \; k \le n \tag{11}$$

of sample data is considered. The empirical cdf, $S_n(x)$, is evaluated for each of the k unique sample values. To evaluate $\hat{F}(x)$, the k unique sample values are standardized by substituting $u = (x-\bar{x})/s$ in $\hat{f}(x)$. The standardized pdf,

$$\hat{\phi}(u) = (1/\sqrt{2\pi}) \exp[-\tfrac{1}{2}u^2], \tag{12}$$

is then used to evaluate $\hat{F}(x)$ which becomes

$$\hat{F}(x) = \hat{\Phi}(u) = \int_{-\infty}^{u} \hat{\phi}(t)dt. \tag{13}$$

In addition to $x_{(1)}$, $x_{(k)}$, $x_{(k)} - x_{(1)}$, n, $\bar{x}$, s^2, and s, system output consists of u, $\hat{F}(x)$, $S_n(x)$, and $|\hat{F}(x) - S_n(x)|$ for each of the k unique order statistics along with the identification of $\hat{D}(n)$. CRT output consists of one or more of the four following plot types.

Plot Type A - $S_n(x)$, $\hat{F}(x)$, F(U), F(L)

Plot Type B - $S_n(x)$, $\hat{F}(x)$, S(U), S(L)

Plot Type C - $S_n(x)$, S(U), S(L)

Plot Type D - $S_n(x)$, $\hat{F}(x)$

The desired plot type(s) are identified on card type 3.

IV. INPUT PREPARATION

A. Deck Setup

The input deck is listed below by card type. Multiple jobs may be processed by stacking card types 1 through 6.

CARD TYPE 1 - JOB IDENTIFICATION CARD

CARD TYPE 2 - VARIABLE FORMAT CARD

CARD TYPE 3 - MAIN CONTROL CARD

CARD TYPE 4 - ALPHA IDENTIFICATION CARD

CARD TYPE 5 - TRANSFORMATION CONSTANT CARD

CARD TYPE 6 - SAMPLE DATA CARD

B. Input Deck Description

Entries for variables with an I format specification must be right-adjusted in the specified field.

CARD TYPE 1 - JOB IDENTIFICATION CARD

Column	Format	Variable	Description
1-80	10A8	JOB(1) - JOB(10)	Job Identification

CARD TYPE 2 - VARIABLE FORMAT CARD

Column	Format	Variable	Description
1-80	10A8	FMT(1) - FMT(10)	Format for reading in the sample data. The format specifications must be enclosed in parentheses.

CARD TYPE 3 - MAIN CONTROL CARD

Column	Format	Variable	Description
1-2	I2	NRUN	Number of transformations to be performed ($1 \leq NRUN \leq 13$)

Column	Format	Variable	Description
11	I1	ITRAN(1)	0 - Do not perform transformation number 1 1 - Do perform transformation number 1
12	I1	ITRAN(2)	0 - Do not perform transformation number 2 1 - Do perform transformation number 2
:	:	:	:
:	:	:	:
:	:	:	:
23	I1	ITRAN(13)	0 - Do not perform transformation number 13 1 - Do perform transformation number 13
25	I2	MEDIUM	2 - Sample data is input on cards 1 - Sample data is input on tape
31	I2	IA	0 - Do not perform plot type A 1 - Do perform plot type A
32	I2	IB	0 - Do not perform plot type B 1 - Do perform plot type B
33	I2	IC	0 - Do not perform plot type C 1 - Do perform plot type C
34	I2	ID	0 - Do not perform plot type D 1 - Do perform plot type D
40-41	I2	NØTIC	Number of desired tick marks on the abscissa axis (if left blank NØTIC is set equal to 15)

Column	Format	Variable	Description
44-45	I2	IRA(1)	Number of decimal digits to be used in labeling abscissa tick marks for transformation number 1
46-47	I2	IRA(2)	Number of decimal digits to be used in labeling abscissa tick marks for transformation number 2
:	:	:	:
:	:	:	:
:	:	:	:
68-69	I2	IRA(13)	Number of decimal digits to be used in labeling abscissa tick marks for transformation number 13
74-75	I2	NBR	Number of sample data per card
78	I1	NC	0 - Card type 5 is not input 1 - Card type 5 is input

CARD TYPE 4 - ALPHA IDENTIFICATION CARD

To obtain plots, at least one alpha must be indicated.

Column	Format	Variable	Description
1	I1	IALPHA(1,1)	1 - Alpha = .20 is desired for transformation number 1 0 - Alpha = .20 is not desired for transformation number 1
2	I1	IALPHA(2,1)	1 - Alpha = .15 is desired for transformation number 1 0 - Alpha = .15 is not desired for transformation number 1

Column	Format	Variable	Description
3	I1	IALPHA(3,1)	1 - Alpha = .10 is desired for transformation number 1 0 - Alpha = .10 is not desired for transformation number 1
4	I1	IALPHA(4,1)	1 - Alpha = .05 is desired for transformation number 1 0 - Alpha = .05 is not desired for transformation number 1
5	I1	IALPHA(5,1)	1 - Alpha = .01 is desired for transformation number 1 0 - Alpha = .01 is not desired for transformation number 1
6	I1	IALPHA(1,2)	1 - Alpha = .20 is desired for transformation number 2 0 - Alpha = .20 is not desired for transformation number 2
:	:	:	:
:	:	:	:
65	I1	IALPHA(5,13)	1 - Alpha = .01 is desired for transformation number 13 0 - Alpha = .01 is not desired for transformation number 13

CARD TYPE 5 - TRANSFORMATION CONSTANT CARD

This card type is omitted if NC = 0 on card type 3.

Column	Format	Variable	Description
1-14	F14.6	A	Constant for transformation number 4
15-28	F14.6	B	Constant for transformation number 5
29-42	F14.6	C	Constant for transformation number 5
43-56	F14.6	D	Constant for transformation number 8
57-70	F14.6	E	Constant for transformation number 11

CARD TYPE 6 - SAMPLE DATA CARD(S)

The sample size (n) must be $4 \leq n \leq 2500$. This card type is omitted if MEDIUM = 1 on card type 3.

Column	Format	Variable	Description
1-2	I2	ITEST	Blank - This is not the last card containing sample data ≠0 - Number of sample data on the last card or record
3-80	Variable	XX	The array of sample data which must be input according to the variable format on card type 2

C. **Request Sheets**

If the sample data (card type 6) is input on purched cards, the job request sheet is prepared as shown below.

SECURITY CLASSIFICATION	NAME	IDENT. NO.	ROOM	BLDG.	PHONE	
☐ TS ☐ S. ☐ C. ☒ U.	J. ORSULAK	A3	A141	1200	7309	SETUP # ___ OF ___

		JOB CARD								JOB TITLE		
☐ COMPILE ☒ GO	☐ CK OUT	CHARGE CODE				IDENT.		PROGRAMMER		K-Test		
☐ COMPILGO	☒ PROD	2	1	5	8	K	8	A	3	EST. COMPILER TIME	EST. EXECUTION TIME 30 SEC.	DATE 12/5/67

TAPES CALLED FOR BY PROBLEM PROGRAM

TAPE NUMBER	Scratch									
FILE PROTECT ON										
PROGRAMMER NUMBER										
SPECIAL HANDLING *(See attached inst.)*										

OPERATOR'S COMMENTS

☐ ABEOJ ☐ HOLD ☐ LOOP ☐ IF CHECKED, SEE REVERSE SIDE FOR ADDITIONAL COMMENTS.

SPECIAL INSTRUCTIONS *(continued on reverse)*

7090 JOB REQUEST NOW-NWL-5230/29 (REV. 11-66)

If the sample data is input on tape, the tape number is written on the job request sheet instead of "Scratch". And the "REEL, NUL" card (second card) of the IOD deck is prepared as follows.

Column

1	Punch the letter "B".
10-17	Punch "REEL, NUL" (or "REEL, PUL" if the tape is file protected).
18-21	Punch the number of the tape.

The CRT request sheet is prepared as follows.

TRAID CAMERA OUTPUT FILM OR PAPER COPIES REQUEST
NDW-NWL-5800 2 (REV. 6-66)

PROGRAMMER	ROOM	BLDG.	PHONE	DATE
J. ORSULAK	A141	1200	7309	12/5/67

FILM IDENTIFICATION
KS-A3

APPROXIMATE NUMBER OF FRAMES
The total number of plots requested for the jobs being processed.

NUMBER OF PAPER COPIES PER FRAME
The number of copies desired of each plot.

FOR OPERATORS ONLY

COUNTER READING START	FINISH

DATE AND TIME PROBLEM RAN

V. OUTPUT FORMAT

A. System Output

(Job identification as given on card type 1)

FORMAT FOR SAMPLE DATA __________

NO. OF RUNS = __   IRUN = __

TRANSFORMATION 1 = __   TRANSFORMATION 2 = __   TRANSFORMATION 3 = __   TRANSFORMATION 4 = __

TRANSFORMATION 5 = __   TRANSFORMATION 6 = __   TRANSFORMATION 7 = __   TRANSFORMATION 8 = __

TRANSFORMATION 9 = __   TRANSFORMATION 10 = __   TRANSFORMATION 11 = __   TRANSFORMATION 12 = __

TRANSFORMATION 13 = __   MEDIUM = __

PLOT TYPE A = __   PLOT TYPE B = __   PLOT TYPE C = __   PLOT TYPE D = __

NOTIC = __   IRA = __ __ __ __ __ __ __ __ __ __ __ __ __

NBR = __   NC - __   ITEST = __

ALPHA IDENTIFICATION = _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____

TRANSFORMATION CONSTANT CARD   A = __________   B = __________   C = __________

D __________   E = __________

ORIGINAL DATA	TRANSFORMED DATA	ORIGINAL DATA	TRANSFORMED DATA
x_1	x_1'	x_2	x_2'
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{i-1}	x_{i-1}'	x_i	x_i'
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{n-1}	x_{n-1}'	x_n	x_n'

MINIMUM = $\underline{x'_{(1)}}$ MAXIMUM = $\underline{x'_{(k)}}$ RANGE = $\underline{x'_{(k)} - x'_{(1)}}$

SAMPLE SIZE = $\underline{n}$ MEAN = $\underline{\bar{x}'}$

VARIANCE = $\underline{s^2}$ STANDARD DEVIATION = $\underline{s}$

KOLMOGOROV TEST FOR NORMALITY

ORDER INDEX (J)	TRANSFORMED DATA X(J)	FREQUENCY	STANDARDIZED VARIATE U(J)	THEORETICAL CDF F(J)	OBSERVED CDF S(J)	ABSOLUTE VALUE F(J)-S(J-1)	ABSOLUTE VALUE F(J)-S(J)
1	$x'_{(1)}$	$f_{(1)}$	$u_{(1)}$	$F_{(1)}$	$S_{(1)}$	$\lvert F_{(1)}-S_{(0)}\rvert$	$\lvert F_{(1)}-S_{(1)}\rvert$
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
j-1	$x'_{(j-1)}$	$f_{(j-1)}$	$u_{(j-1)}$	$F_{(j-1)}$	$S_{(j-1)}$	$\lvert F_{(j-1)}-S_{(j-2)}\rvert$	$\lvert F_{(j-1)}-S_{(j-1)}\rvert$
j	$x'_{(j)}$	$f_{(j)}$	$u_{(j)}$	$F_{(j)}$	$S_{(j)}$	$\lvert F_{(j)}-S_{(j-1)}\rvert$	$\lvert F_{(j)}-S_{(j)}\rvert$
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
k	$x'_{(k)}$	$f_{(k)}$	$u_{(k)}$	$F_{(k)}$	$S_{(k)}=1$	$\lvert F_{(k)}-S_{(k-1)}\rvert$	$\lvert F_{(k)}-S_{(k)}\rvert$

D(ORDER INDEX) = $\underline{\hat{D}(n)}$

Note: $f_{(\cdot)} = \hat{f}(x'_{(j)})$, $F_{(\cdot)} = \hat{F}(x'_{(j)})$, and $S_{(\cdot)} = S_n(x'_{(j)})$

B. CRT Output

The plots are performed by the system "GRF Plot Subroutine" and are plotted on the CRT printer units. The plots are in an ordinary Cartesian coordinate system with origin $(x'_{(1)}, 0)$. The abscissa of the system is the x'-axis. A zero (0) is plotted at the point $(\bar{x}', 0.5)$.

The four types of plots are labeled as follows.

Plot Type A - $S_n(x)$, $\hat{F}(x)$, F(U), F(L)

OBSERVED RELATIVE CUMULATIVE FREQUENCY,
THEORETICAL RELATIVE CUMULATIVE FREQUENCY, AND
CRITICAL REGION BOUNDS FOR ALPHA = α LEVEL OF SIGNIFICANCE,
D(ALPHA) = $D_\alpha(n)$

Plot Type B - $S_n(x)$, $\hat{F}(x)$, S(U), S(L)

OBSERVED RELATIVE CUMULATIVE FREQUENCY,
THEORETICAL RELATIVE CUMULATIVE FREQUENCY,
$100(1-\alpha)$ PERCENT CONFIDENCE BANDS FOR THE THEORETICAL
CUMULATIVE DISTRIBUTION FUNCTION,
D(ALPHA) = $D_\alpha(n)$

Plot Type C - $S_n(x)$, S(U), S(L)

OBSERVED RELATIVE CUMULATIVE FREQUENCY AND THE
$100(1-\alpha)$ PERCENT CONFIDENCE BANDS FOR THE THEORETICAL
CUMULATIVE DISTRIBUTION FUNCTION

Plot Type D - $S_n(x)$, $\hat{F}(x)$

OBSERVED AND THEORETICAL RELATIVE CUMULATIVE FREQUENCY

C. Program Running Time

The running time is printed at the end of each job processed. The following table of observed times may be used as a guide for estimating program running times.

Sample Size	No. of Runs	Total No. of Plots	Time in Seconds
25	1	4	12.4
140	1	4	16.1
10	1	6	19.0
30	1	6	20.0
20	2	8	25.4
20	4	16	45.8
20	5	20	56.4
1000	1	2	54.6

VI. EXAMPLE PROBLEM

A. Problem Description

The example problem consists of 20 observations (x_i -values) randomly sampled from a continuous distribution. Two runs were performed, one on the original observations and one on the logarithmically transformed values. Because the minimum original sample value was -750, transformation number 4 was used with the constant A = 751. All four available plot types were performed. The 0.05-level of significance was chosen for testing the null hypothesis of normality.

The input deck setup is illustrated on the following Data Card Layout Sheet.

B. Input for Example Problem

7090 DATA CARD LAYOUT
NUMBER ONE
NOB-NRL-10882 60 (REV 3-64)

PAGE 1 of 1

DATA CARD LAYOUT — PROG IDEN. — CARD NUMBER

```
  EXAMPLE PROBLEM
(I2,7F10.5)
 2        1  1          2     1111      15  0     2                       7  1
   1              1
     751.
     -743.      130.     -130.     -424.     -743.     -742.     -699.
     -718.     -730.     -745.     -750.      775.     -378.     -378.
 6     -8.     -733.     -746.     -739.     -490.     -624.
```

C. Program Output

```
  EXAMPLE PROBLEM
FORMAT FOR SAMPLE DATA      (I2,7F10.5)
NO. OF RUNS =  2     IRUN =  1
TRANSFORMATION  1 = 1     TRANSFORMATION  2 = 0     TRANSFORMATION  3 = 0     TRANSFORMATION  4 = 1
TRANSFORMATION  5 = 0     TRANSFORMATION  6 = 0     TRANSFORMATION  7 = 0     TRANSFORMATION  8 = 0
TRANSFORMATION  9 = 0     TRANSFORMATION 10 = 0     TRANSFORMATION 11 = 0     TRANSFORMATION 12 = 0
TRANSFORMATION 13 = 0     MEDIUM = 2
PLOT TYPE A = 1           PLOT TYPE B = 1           PLOT TYPE C = 1           PLOT TYPE D = 1
NOTIC = 15                IRA =  0  0  0  2  0  0  0  0  0  0  0  0  0
NBR =  7                  NC = 1                    ITEST =  6
ALPHA IDENTIFICATION = 00010 00000 00000 00010 00000 00000 00000 00000 00000 00000 00000 00000 00000
TRANSFORMATION CONSTANT CARD   A =    751.000000   B =     -.000000   C =     -.000000
                               D =     -.000000    E =     -.000000
```

ORIGINAL DATA	TRANSFORMED DATA	ORIGINAL DATA	TRANSFORMED DATA
-743.0000000000	-743.0000000000	130.0000000000	130.0000000000
-130.0000000000	-130.0000000000	-424.0000000000	-424.0000000000
-743.0000000000	-743.0000000000	-742.0000000000	-742.0000000000
-699.0000000000	-699.0000000000	-718.0000000000	-718.0000000000
-730.0000000000	-730.0000000000	-745.0000000000	-745.0000000000
-750.0000000000	-750.0000000000	775.0000000000	775.0000000000
-378.0000000000	-378.0000000000	-378.0000000000	-378.0000000000
-8.0000000000	-8.0000000000	-733.0000000000	-733.0000000000
-746.0000000000	-746.0000000000	-739.0000000000	-739.0000000000
-490.0000000000	-490.0000000000	-624.0000000000	-624.0000000000

MINIMUM = -.75000000E+03
SAMPLE SIZE = 20
VARIANCE = .16085451E+06

MAXIMUM = .77500000E+03
MEAN = -.48075000E+03
STANDARD DEVIATION = .40106672E+03

RANGE = .15250000E+04

KOLMOGOROV TEST FOR NORMALITY

ORDER INDEX (J)	TRANSFORMED DATA X(J)	FREQUENCY	STANDARDIZED VARIATE U(J)	THEORETICAL CDF F(J)	OBSERVED CDF S(J)	ABSOLUTE VALUE F(J)-S(J-1)	ABSOLUTE VALUE F(J)-S(J)
1	-750.000000	1	-.67133468619	0.251005376963	0.050000000000	.2510053769633	.2010053769633
2	-746.000000	1	-.66136128324	0.254191995335	0.100000000000	.2041919953354	.1541919953354
3	-745.000000	1	-.65886793250	0.254991951900	0.150000000000	.1549919519005	.1049919519005
4	-743.000000	2	-.65388123103	0.256595807833	0.250000000000	.1065958078329	.0065958078329
5	-742.000000	1	-.65138788029	0.257399701526	0.300000000000	.0073997015265	.0426002984735
6	-739.000000	1	-.64390782808	0.259819211035	0.350000000000	.0401807889647	.0901807889647
7	-733.000000	1	-.62894772365	0.264693193728	0.400000000000	.0853068062717	.1353068062717
8	-730.000000	1	-.62146767143	0.267147504709	0.450000000000	.1328524952909	.1828524952909
9	-718.000000	1	-.59154746258	0.277078242255	0.500000000000	.1729217577448	.2229217577446
10	-699.000000	1	-.54417379856	0.293162163623	0.550000000000	.2068378363766	.2568378363766
11	-624.000000	1	-.35717249321	0.360481896148	0.600000000000	.1895181038518	.2395181038518
12	-490.000000	1	-.02306349433	0.490799882937	0.650000000000	.1092001170631	.15920011706[illegible]
13	-424.000000	1	0.14149765438	0.556261363036	0.700000000000	.0937386369636	.1437386369
14	-378.000000	2	0.25619178832	0.601098282164	0.800000000000	.0989017178364	.198901717[illegible]
15	-130.000000	1	0.87454277134	0.809086423934	0.850000000000	.0090864239343	.040913576065[illegible]
16	-8.000000	1	1.17873156137	0.880745227933	0.900000000000	.0307452279333	.0192547720667
17	130.000000	1	1.52281396320	0.936095490092	0.950000000000	.0360954900925	.0139045099075
18	775.000000	1	3.13102518918	0.999128962704	1.000000000000	.0491289627045	.0008710372955

D(10) = 0.2568

RANGE OVERFLOW *

RANGE OVERFLOW

RANGE OVERFLOW

RANGE OVERFLOW

RANGE OVERFLOW

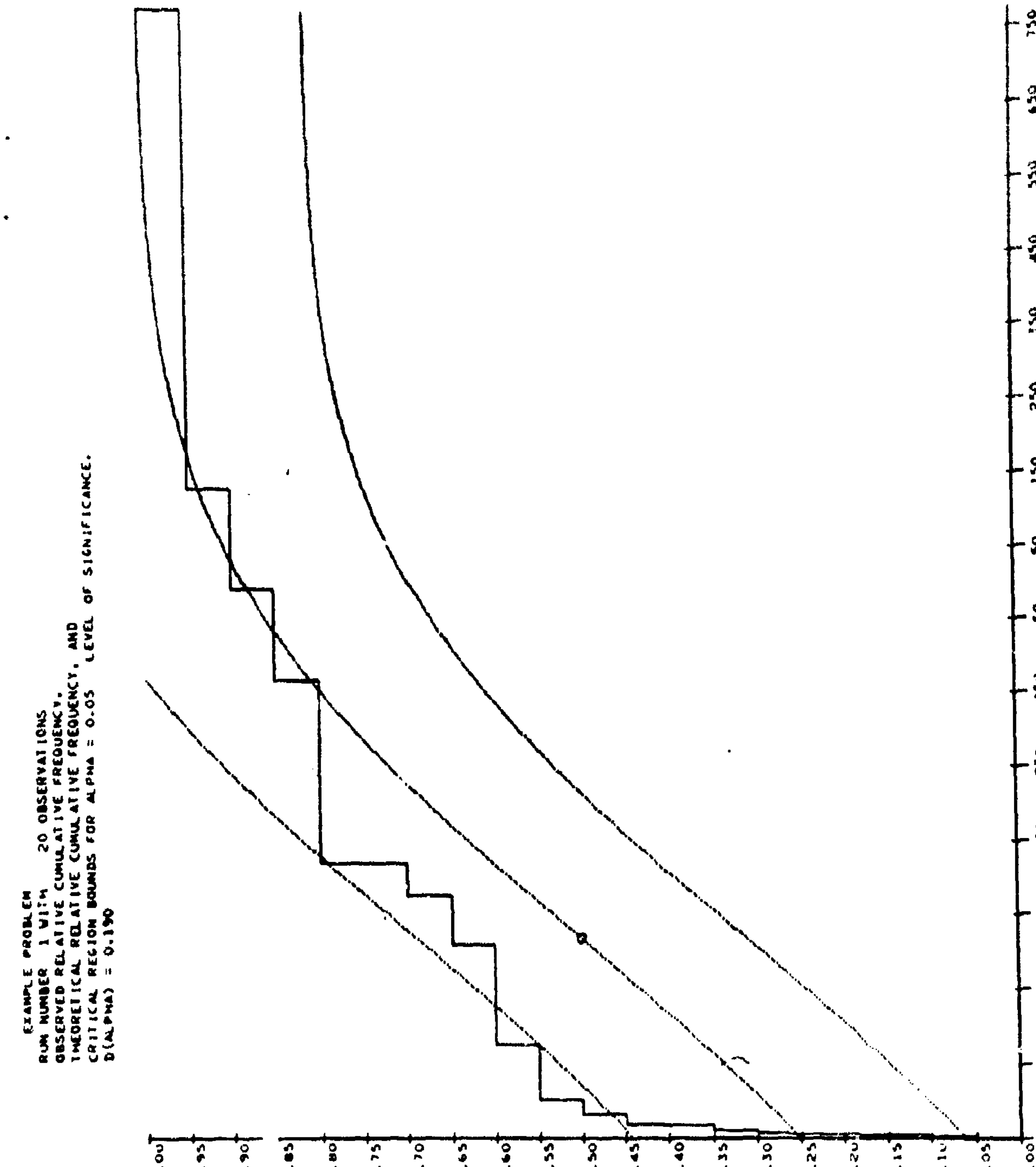
RANGE OVERFLOW

RANGE OVERFLOW

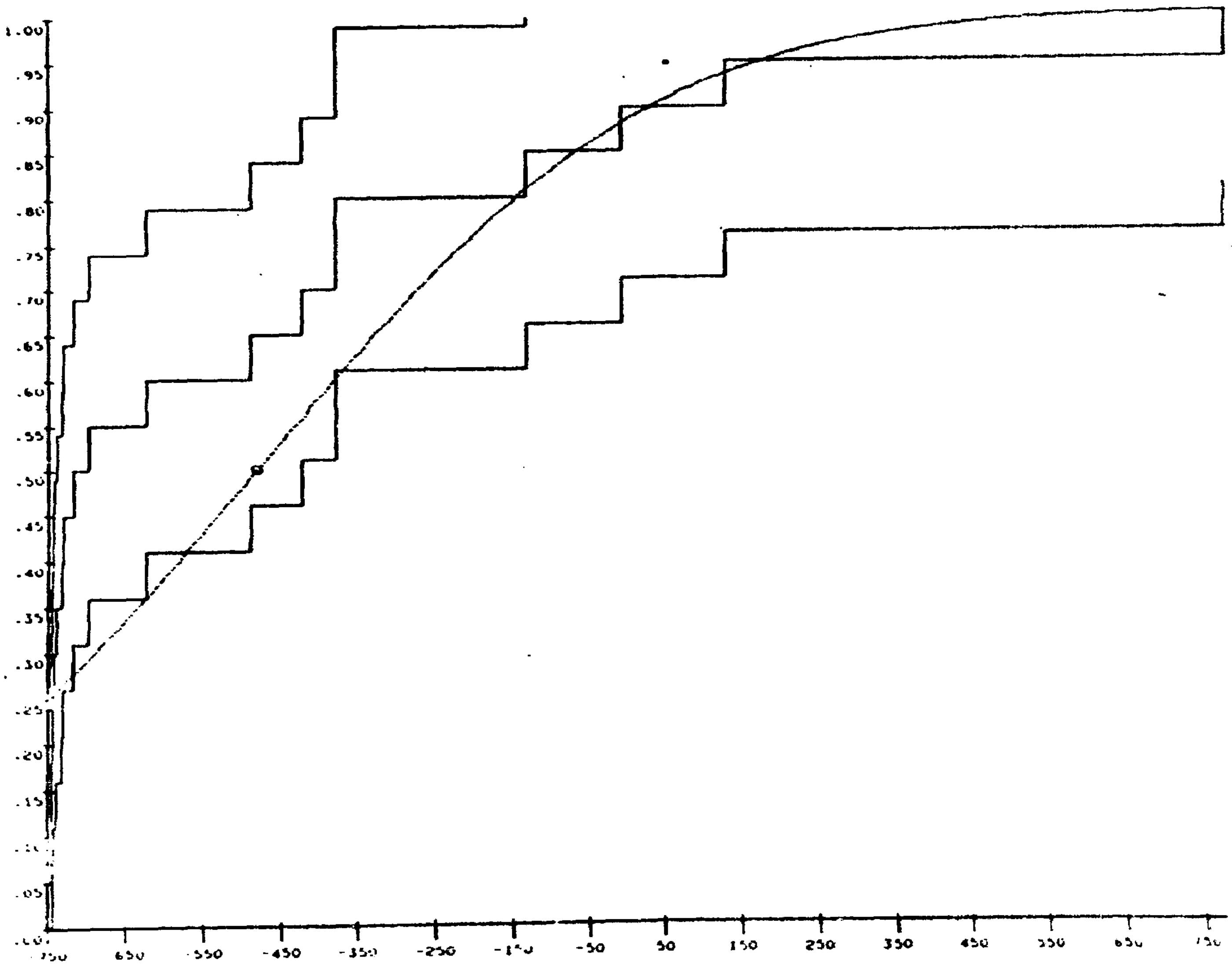
RANGE OVERFLOW

RANGE OVERFLOW

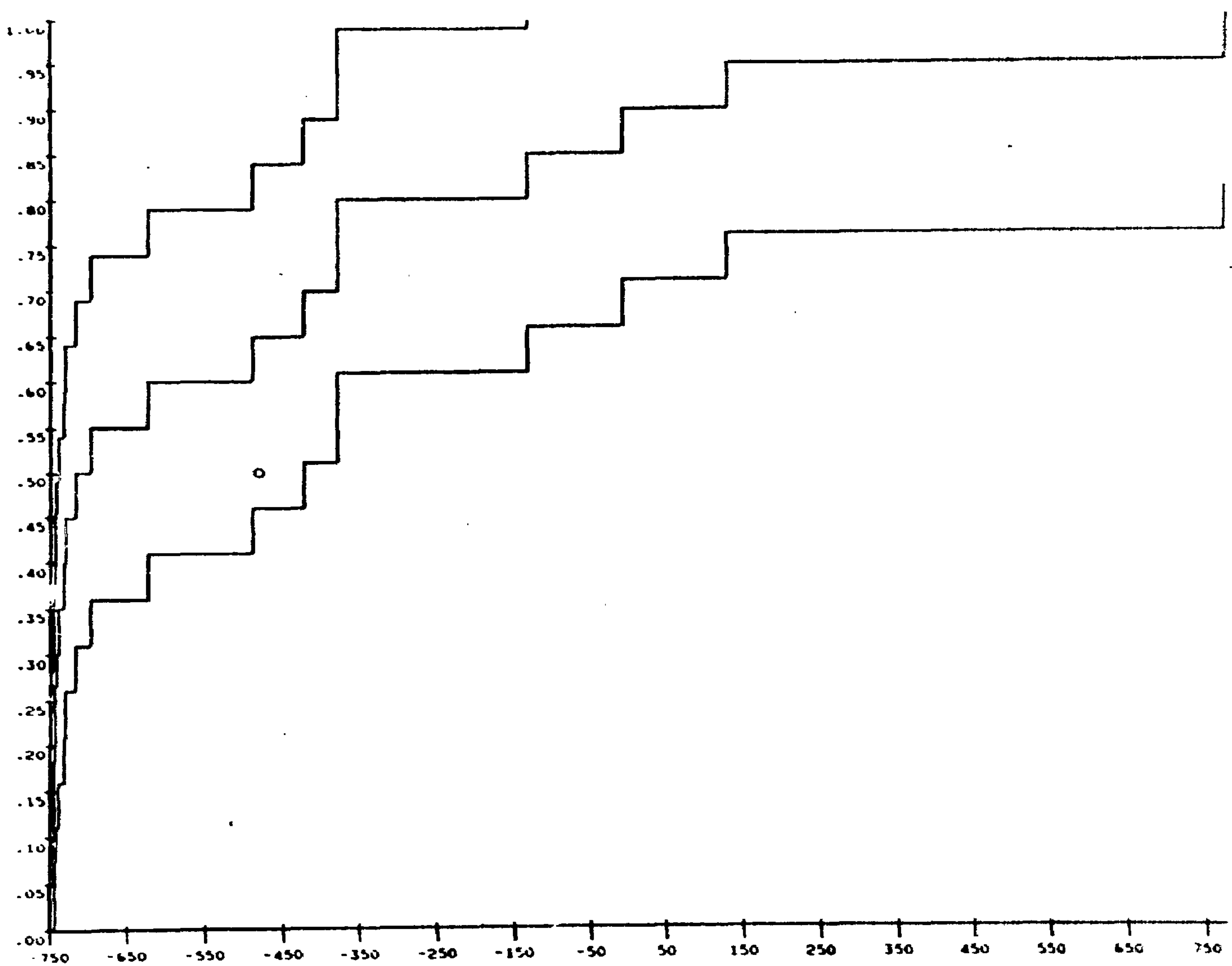
*RANGE OVERFLOW indicates one or more points are not within the ranges of the axes.


EXAMPLE PROBLEM
RUN NUMBER 1 WITH 20 OBSERVATIONS
OBSERVED RELATIVE CUMULATIVE FREQUENCY,
THEORETICAL RELATIVE CUMULATIVE FREQUENCY, AND
CRITICAL REGION BOUNDS FOR ALPHA = 0.05 LEVEL OF SIGNIFICANCE.
D(ALPHA) = 0.190
1.00
.95
.90
.85
.80
.75
.70
.65
.60
.55
.50
.45
.40
.35
.30
.25
.20
.15
.10
.05
.00
-750
-650
-550
-450
-350
-250
-150
-50
50
150
250
350
450
550
650
750

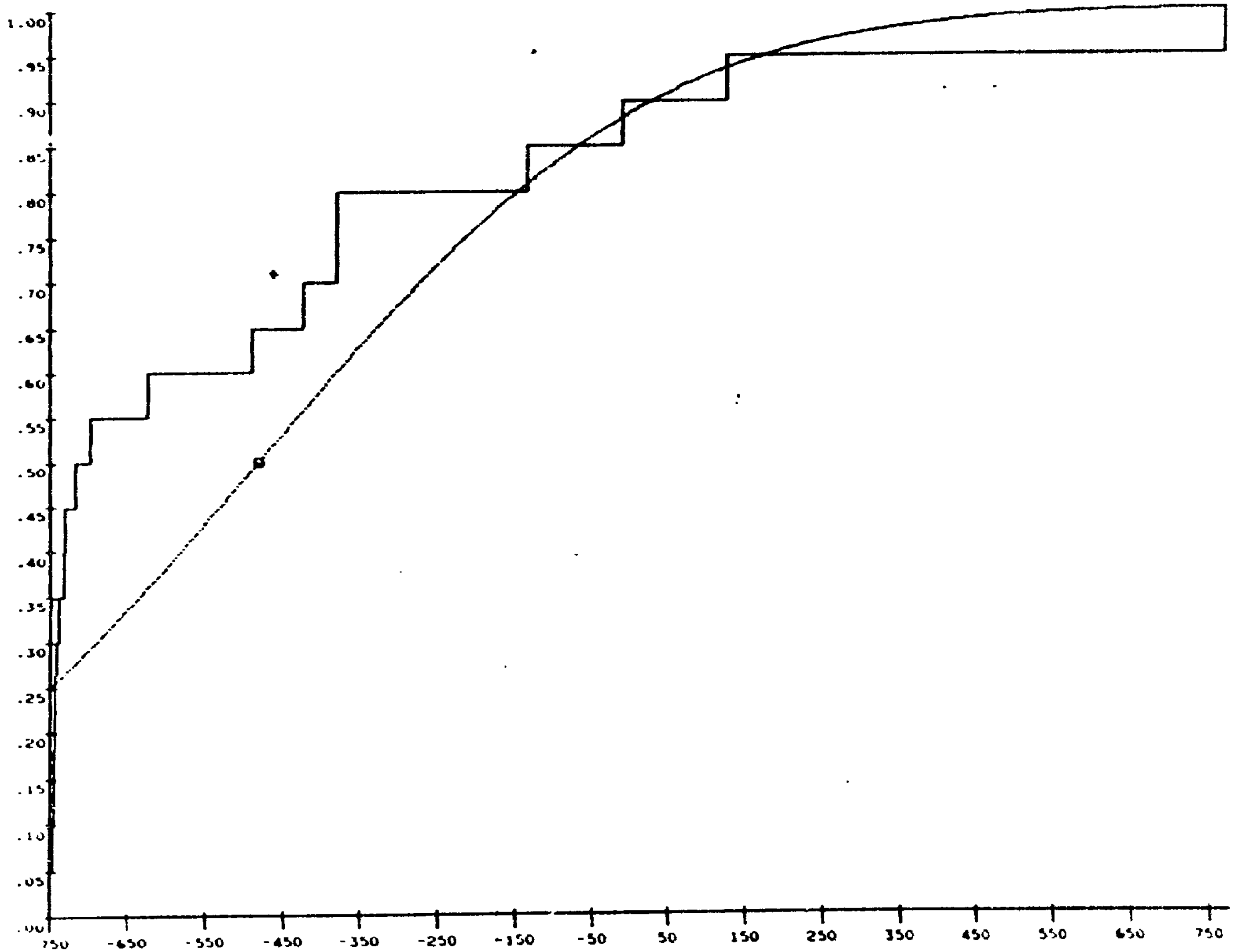
EXAMPLE PROBLEM
RUN NUMBER 1 WITH 20 OBSERVATIONS
OBSERVED RELATIVE CUMULATIVE FREQUENCY.
THEORETICAL RELATIVE CUMULATIVE FREQUENCY.
95 PERCENT CONFIDENCE BANDS FOR THE THEORETICAL CUMULATIVE DISTRIBUTION FUNCTION.
D(ALPHA) = 0.190



EXAMPLE PROBLEM
RUN NUMBER 1 WITH 20 OBSERVATIONS
OBSERVED RELATIVE CUMULATIVE FREQUENCY AND THE
95 PERCENT CONFIDENCE BANDS FOR THE THEORETICAL CUMULATIVE DISTRIBUTION FUNCTION



EXAMPLE PROBLEM
RUN NUMBER 1 WITH 20 OBSERVATIONS
OBSERVED AND THEORETICAL RELATIVE CUMULATIVE FREQUENCY



```
  EXAMPLE PROBLEM
FORMAT FOR SAMPLE DATA      (I2,7F10.5)
NO. OF RUNS =  2      IRUN =  2
TRANSFORMATION  1 = 1      TRANSFORMATION  2 = 0      TRANSFORMATION  3 = 0      TRANSFORMATION  4 = 1
TRANSFORMATION  5 = 0      TRANSFORMATION  6 = 0      TRANSFORMATION  7 = 0      TRANSFORMATION  8 = 0
TRANSFORMATION  9 = 0      TRANSFORMATION 10 = 0      TRANSFORMATION 11 = 0      TRANSFORMATION 12 = 0
TRANSFORMATION 13 = 0      MEDIUM = 2
PLOT TYPE A = 1            PLOT TYPE B = 1            PLOT TYPE C = 1            PLOT TYPE D = 1
NOTIC = 15                 IRA =  0  0  0  2  0  0  0  0  0  0  0  0  0
NBR =  7                   NC = 1                     ITEST =  6
ALPHA IDENTIFICATION = 00010 00000 00000 00010 00000 00000 00000 00000 00000 00000 00000 00000 00000
TRANSFORMATION CONSTANT CARD   A =      751.000000   B =        -.000000   C =        -.000000
                               D =        -.000000   E =        -.000000
```

ORIGINAL DATA	TRANSFORMED DATA	ORIGINAL DATA	TRANSFORMED DATA
-743.0000000000	2.0794415417	130.0000000000	6.7810576259
-130.0000000000	6.4313310819	-424.0000000000	5.7899601709
-743.0000000000	2.0794415417	-742.0000000000	2.1972245773
-699.0000000000	3.9512437186	-718.0000000000	3.4965075615
-730.0000000000	3.0445224377	-745.0000000000	1.7917594692
-750.0000000000	0.0000000000	775.0000000000	7.3304052118
-378.0000000000	5.9215784196	-378.0000000000	5.9215784196
-8.0000000000	6.6106960447	-733.0000000000	2.8903717579
-746.0000000000	1.6094379124	-739.0000000000	2.4849066498
-490.0000000000	5.5645204073	-624.0000000000	4.8441870865

MINIMUM =	.43231173E-14	MAXIMUM =	.73304052E+01	RANGE =	.73304052E+01
SAMPLE SIZE =	20	MEAN =	.40410086E+01		
VARIANCE =	.45885482E+01	STANDARD DEVIATION =	.21420897E+01		

KOLMOGOROV TEST FOR NORMALITY

ORDER INDEX (J)	TRANSFORMED DATA X(J)	FREQUENCY	STANDARDIZED VARIATE U(J)	THEORETICAL CDF F(J)	OBSERVED CDF S(J)	ABSOLUTE VALUE F(J)-S(J-1)	ABSOLUTE VALUE F(J)-S(J)
1	0.000000	1	-1.88647964775	0.029616622275	0.050000000000	.0296166222748	.0203833777252
2	1.609438	1	-1.13513952940	0.128160693262	0.100000000000	.0781606932623	.0281606932623
3	1.791759	1	-1.05002565268	0.146855459970	0.150000000000	.0468554599698	.0031445400302
4	2.079442	2	-.91572591940	0.179907588464	0.250000000000	.0299075884644	.0700924115356
5	2.197225	1	-.86074080984	0.194692592440	0.300000000000	.0553074075600	.1053074075600
6	2.484907	1	-.72644107657	0.233786091575	0.350000000000	.0662139084254	.1162139084254
7	2.890372	1	-.53715623373	0.295581045559	0.400000000000	.0544189544409	.1044189544409
8	3.044522	1	-.46519347634	0.320897370431	0.450000000000	.0791026295686	.1291026295686
9	3.496508	1	-.25419151488	0.399674143108	0.500000000000	.0503258568921	.1003258568921
10	3.951244	1	-.04190527789	0.483287218576	0.550000000000	.0167127814236	.0667127814236
11	4.844187	1	0.37495092422	0.646150908784	0.600000000000	.0961509087841	.0461509087841
12	5.564520	1	0.71122691124	0.761526345987	0.650000000000	.1615263459866	.1115263459866
13	5.789960	1	0.81646982700	0.792882154847	0.700000000000	.1428821548473	.0928821548473
14	5.921578	2	0.87791368254	0.810002506731	0.800000000000	.1100025067313	.0100025067313
15	6.431331	1	1.11588348719	0.867761720013	0.850000000000	.0677617200130	.0177617200130
16	6.610696	1	1.19961712570	0.884853751445	0.900000000000	.0348537514454	.0151462485546
17	6.781058	1	1.27914768081	0.899575335841	0.950000000000	.0004246641593	.0504246641593
18	7.330405	1	1.53560173662	0.937680087110	1.000000000000	.0123199128895	.0623199128895

D(12) = 0.1615

RANGE OVERFLOW

RANGE OVERFLOW

RANGE OVERFLOW

RANGE OVERFLOW

RANGE OVERFLOW

RANGE OVERFLOW

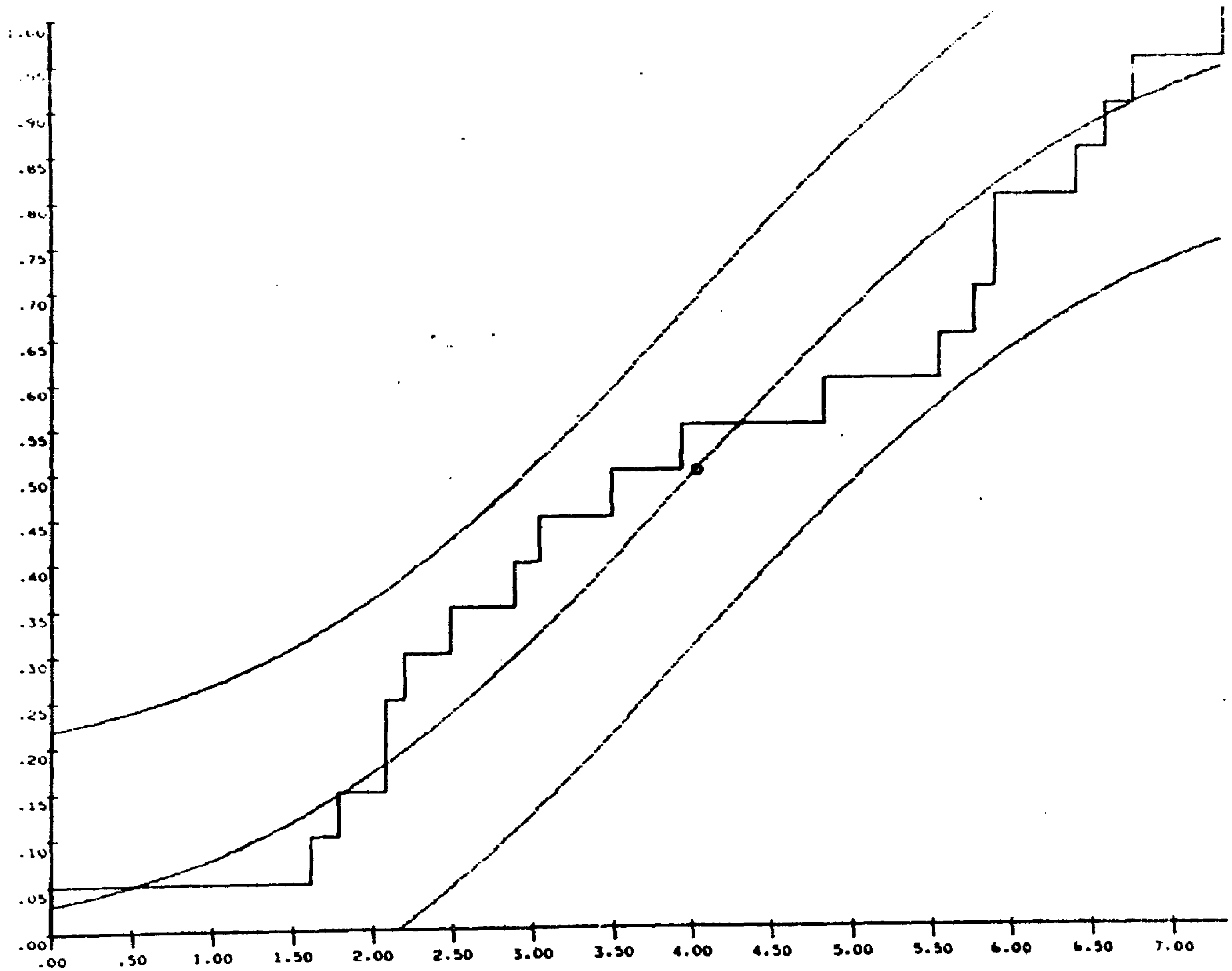
RANGE OVERFLOW

RANGE OVERFLOW

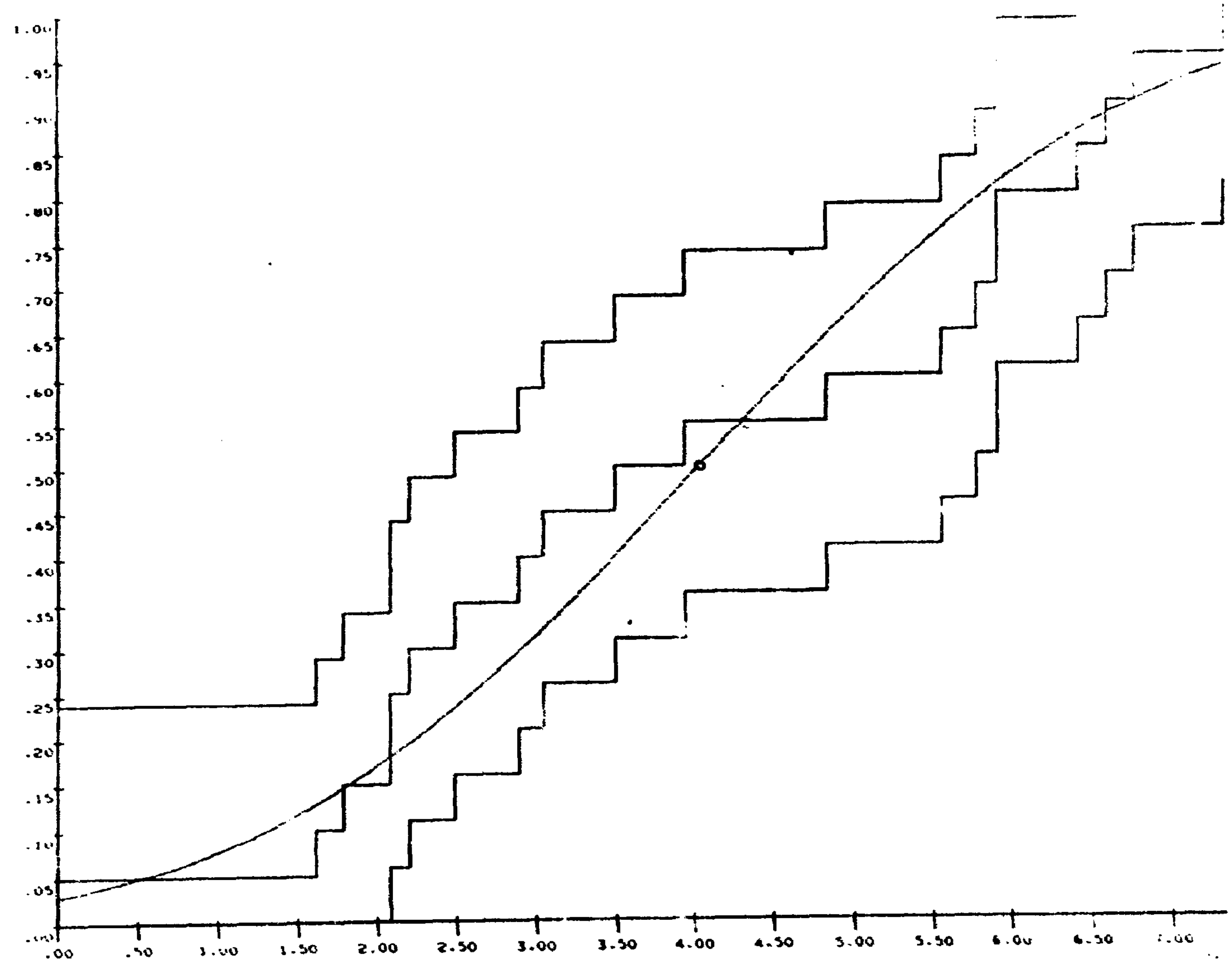
RANGE OVERFLOW

TIME IN SECONDS = 25.39

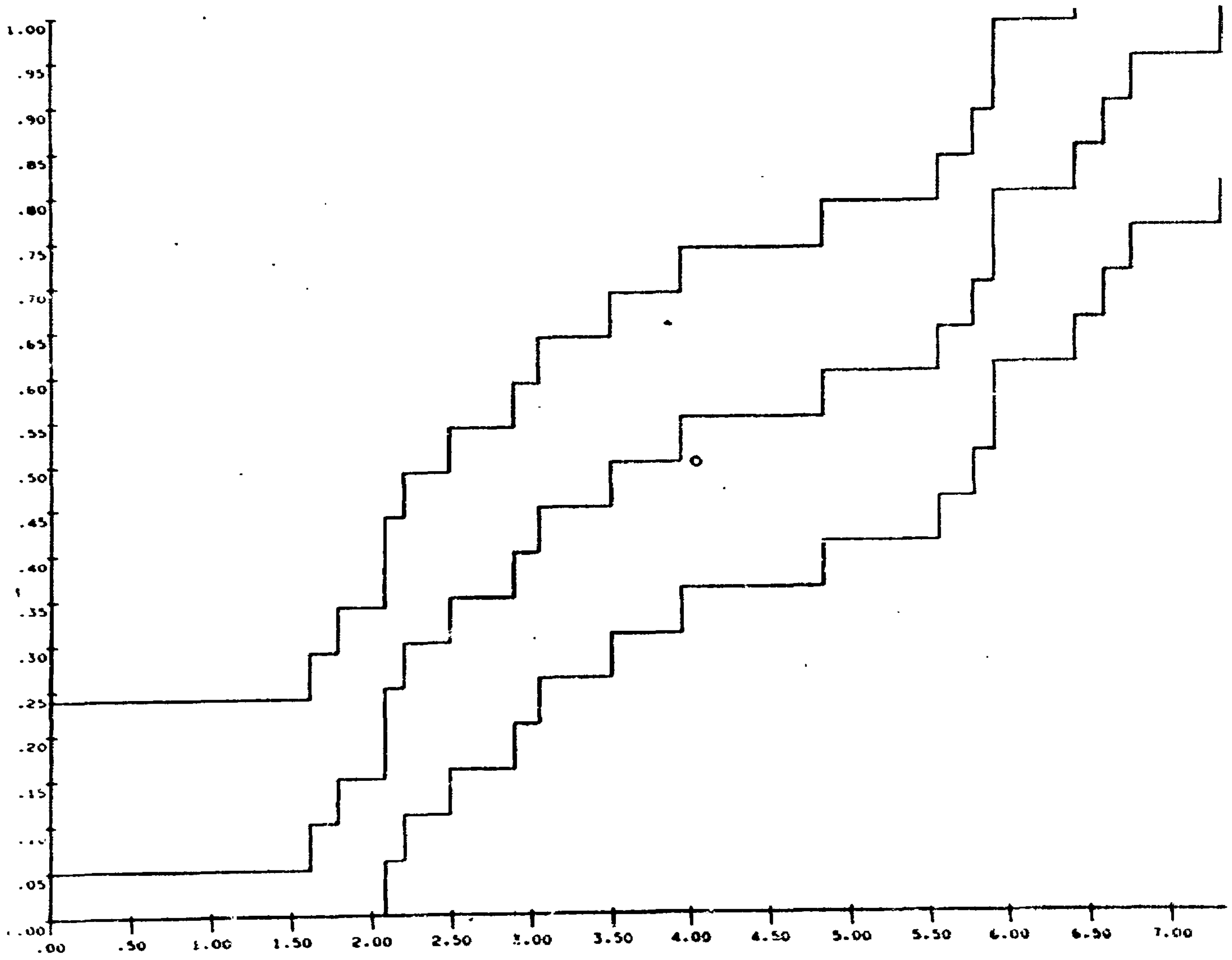
EXAMPLE PROBLEM
RUN NUMBER 2 WITH 20 OBSERVATIONS
OBSERVED RELATIVE CUMULATIVE FREQUENCY,
THEORETICAL RELATIVE CUMULATIVE FREQUENCY, AND
CRITICAL REGION BOUNDS FOR ALPHA = 0.05 LEVEL OF SIGNIFICANCE.
D(ALPHA) = 0.190



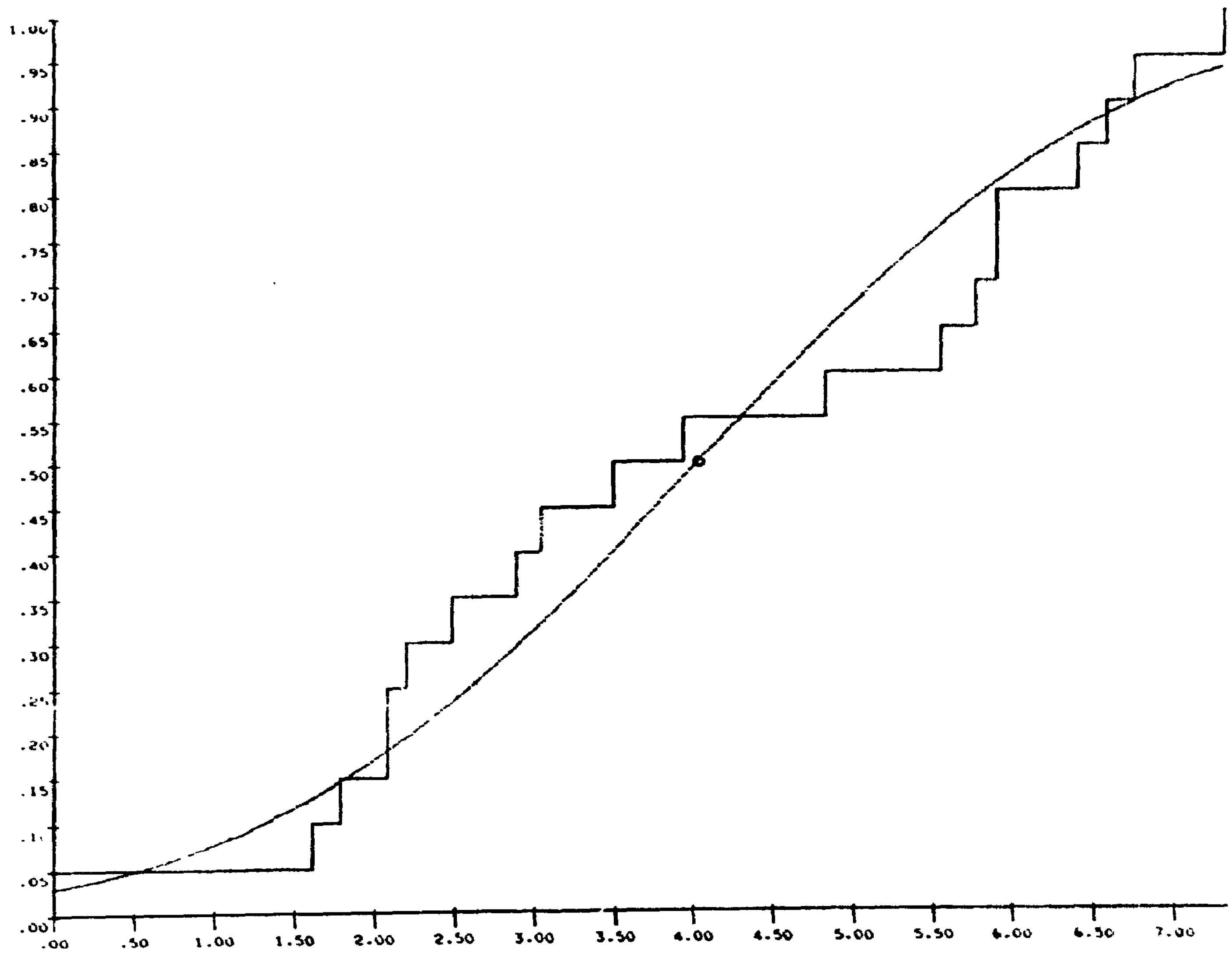
EXAMPLE PROBLEM
RUN NUMBER 2 WITH 20 OBSERVATIONS
OBSERVED RELATIVE CUMULATIVE FREQUENCY.
THEORETICAL RELATIVE CUMULATIVE FREQUENCY.
95 PERCENT CONFIDENCE BANDS FOR THE THEORETICAL CUMULATIVE DISTRIBUTION FUNCTION.
D(ALPHA) = 0.190



EXAMPLE PROBLEM
RUN NUMBER 2 WITH 20 OBSERVATIONS
OBSERVED RELATIVE CUMULATIVE FREQUENCY AND THE
95 PERCENT CONFIDENCE BANDS FOR THE THEORETICAL CUMULATIVE DISTRIBUTION FUNCTION



EXAMPLE PROBLEM
RUN NUMBER 2 WITH 20 OBSERVATIONS
OBSERVED AND THEORETICAL RELATIVE CUMULATIVE FREQUENCY



D. Discussion of Test Results

To test the null hypothesis "The 20 randomly sampled observations are from a parent normal population" at the 0.05-level of significance, $\hat{D}(n) = 0.2568$ (from the first run) is compared with $D_{0.05}(20) = 0.190$. The null hypothesis is, therefore, rejected at the 0.05-level of significance. Note that the maximum absolute deviation appears at $x_{(10)} = -699$, and at this point $S_n(x)$ has passed above $F(U)$ into the upper rejection region (see pages 24 and 25). To test the null hypothesis of normality of the transformed random variable at the 0.05-level of significance, $\hat{D}(n) = 0.1615$ (from the second run) is compared with $D_{0.05}(20) = 0.190$ (see page 31); consequently, the null hypothesis is not rejected. Unlike run number 1, plot type A for run number 2 on page 32 shows $S_n(x)$ remaining within the "acceptance" band. Therefore, we conclude at the 0.05-level of significance that the original random variable is approximately log normally distributed.

VII. PROGRAM LISTING

```
B         TYPE,COMPILGO,FORTRAN,PM
T         SUBTYPE,FIOD
B1D       IOD,TAPE,,,,EVEN,,SAVE
B         REEL,NULXXXX
B   2     IOD,BREADER
          END
T         SUBTYPE,FORTRAN,LMAP,PUNCH
C      KOLMOGOROV TEST FOR NORMALITY
       DIMENSION  TABLE(28, 5)
       DIMENSION JOB(10), FMT(10),  IFILE(35), ITRAN(13), IALPHA (5,13),
     1 X(2500), U(2500), ALPHA(5), IFREQ(2500), TEN(20), T(2501), IRA(13
     2 ), XX(2500), S(2500), SN1(2500), SN(2500), F(2500), FN(500),
     3 FL(500), FU(500), SL(2500), SU(2500)
       COMMON /PRIME/ XP(2, 1250), XXP(2, 1250)
       EQUIVALENCE (IFILE(1), NRUN), (IFILE(2), ITRAN(1)),
     1 (IFILE(3), ITRAN(2)), (IFILE(4), ITRAN(3)), (IFILE(5), ITRAN(4)),
     2 (IFILE(6), ITRAN(5)), (IFILE(7), ITRAN(6)), (IFILE(8), ITRAN(7)),
     3 (IFILE(9), ITRAN(8)), (IFILE(10), ITRAN(9)), (IFILE(11), ITRAN(10
     4 )), (IFILE(12), ITRAN(11)), (IFILE(13), ITRAN(12)), (IFILE(14),
     5 ITRAN(13)), (IFILE(15), MEDIUM), (IFILE(16), IA), (IFILE(17), IB)
     6 , (IFILE(18), IC),  (IFILE(19), ID), (IFILE(20), NOTIC)
       EQUIVALENCE (IFILE(21), IRA(1)), (IFILE(22), IRA(2)), (IFILE(23),
     1IRA(3)), (IFILE(24), IRA(4)), (IFILE(25), IRA(5)), (IFILE(26),
     2IRA(6)), (IFILE(27), IRA(7)), (IFILE(28), IRA(8)), (IFILE(29),
     3IRA(9)), (IFILE(30), IRA(10)), (IFILE(31), IRA(11)), (IFILE(32),
     4IRA(12)), (IFILE(33), IRA(13)), (IFILE(34), NBR), (IFILE(35), NC)
       EQUIVALENCE (XP(1,1), X(1)), (XXP(1,1), XX(1))
       DATA (ALPHA(I), I = 1, 5) (.20, .15, .10, .05, .01)
       DATA (TABLE(I,1), I = 1, 28)
     1(.300, .285, .265, .247, .233, .223, .215, .206, .199, .190, .183,
     2 .177, .173, .169, .166, .163, .160, .160, .160, .160, .160, .142,
     3 .142, .142, .142, .142, .131, .736)
       DATA (TABLE (I,2), I = 1, 28)
     1(.319, .299, .277, .258, .244, .233, .224, .217, .212, .202, .194,
     2 .187, .182, .177, .173, .169, .166, .166, .166, .166, .166, .147,
     3 .147, .147, .147, .147, .136, .768)
       DATA (TABLE (I,3), I = 1, 28)
     1(.352, .315, .294, .276, .261, .249, .239, .230, .223, .214, .207,
     2 .201, .195, .189, .184, .179, .174, .174, .174, .174, .174, .158,
     3 .158, .158, .158, .158, .144, .805)
       DATA (TABLE (I,4), I = 1, 28)
     1(.381, .337, .319, .300, .285, .271, .258, .249, .242, .234, .227,
     2 .220, .213, .206, .200, .195, .190, .190, .190, .190, .190, .173,
     3 .173, .173, .173, .173, .161, .886)
       DATA (TABLE (I,5), I = 1, 28)
     1(.417, .405, .364, .348, .331, .311, .294, .284, .275, .268, .261,
     2 .257, .250, .245, .239, .235, .231, .231, .231, .231, .231, .200,
     3 .200, .200, .200, .200, .187, 1.031)
       CALL CRTID (4HKSA3)
       CALL GRSI (.06652, .06652, 1., .06652)
       CALL SETIT
       CALL MOVE (0., XP, 5000)
C      READ INPUT
   100 READ 110, JOB, FMT
   110 FORMAT (10A8/ 10A8)
       READ 120, IFILE
   120 FORMAT (I2, 8X, 13I1, 1X, I1, 5X, 4I1, 5X, I2, 2X, 13I2, 4X, I2,
     12X, I1)
       READ 130, IALPHA
```

```
  130 FORMAT (65I1)
      IF (NC) 140, 160, 140
  140 READ 150, A, B, C, D, E
  150 FORMAT (5F14.6)
  160 K = 1
      KK = NBR
  170 READ (MEDIUM, FMT) ITEST, (XX(I), I = K,KK)
  180 IF (ITEST) 200, 190, 200
  190 K = K + NBR
      KK = KK + NBR
      GO TO 170
  200 N = K + ITEST -1
  210 IRUN = 1
C     TRANSFORMATION 1
      IF (ITRAN(1)) 220, 240, 220
  220   DO 230 I = 1, N
  230    X(I) = XX(I)
        IT = 1
         ASSIGN 240 TO ILINE
         GO TO 740
C     TRANSFORMATION 2
  240 IF (ITRAN(2)) 250, 290, 250
  250   IT = 2
        ASSIGN 290 TO ILINE
        DO  280 I = 1, N
        IF(XX(I)) 260, 280, 280
  260   PRINT 270, IT
  270   FORMAT (24H ERROR IN TRANSFORMATION, 3X, I2, 3X,21HLN OF NEGATIV
     1E NUMBER)
        GO TO 1330
  280   X(I) = ALOG(XX(I))
        GO TO 740
C     TRANSFORMATION 3
  290 IF (ITRAN(3)) 300, 330, 300
  300   IT = 3
        ASSIGN 330 TO  ILINE
        DO  320 I = 1, N
        IF (XX(I)) 310, 320, 320
  310   PRINT 270, IT
        GO TO 1330
  320   X(I) = ALOG( ALOG(XX(I)))
        GO TO 740
C     TRANSFORMATION 4
  330 IF (ITRAN(4)) 340, 370, 340
  340   IT = 4
        ASSIGN 370 TO ILINE
        DO  360 I = 1, N
        ARG = A + XX(I)
        IF (ARG) 350, 360, 360
  350   PRINT 270, IT
        GO TO 1330
  360   X(I) = ALOG (ARG)
        GO TO 740
C     TRANSFORMATION 5
  370 IF (ITRAN(5)) 380, 420, 380
  380   IT = 5
        ASSIGN 420 TO ILINE
        DO  400 I = 1, N
        ARG = C + XX(I)
```

```
      IF (ARG) 410, 390, 390
  390 ARG2 =  B + ALOG(ARG)
      IF (ARG2) 410, 400, 400
  400 X(I) = ALOG (ARG2)
      GO TO 740
  410 PRINT 270, IT
      GO TO 1330
C     TRANSFORMATION 6
  420 IF (ITRAN(6)) 430, 470, 430
  430 IT = 6
      ASSIGN 470 TO ILINE
      DO  460 I = 1, N
      IF (XX(I)) 440, 460, 460
  440 PRINT 450, IT
  450 FORMAT (24H ERROR IN TRANSFORMATION, 3X, I2, 3X, 32HSQUARE ROOT
     1OF A NEGATIVE NUMBER)
      GO TO 1330
  460 X(I) = SQRT(XX(I))
      GO TO 740
C     TRANSFORMATION 7
  470 IF (ITRAN(7)) 480, 520, 480
  480 IT = 7
      ASSIGN 520 TO ILINE
      DO  510 I = 1, N
      IF (XX(I)) 510, 490, 510
  490 PRINT 500, IT
  500 FORMAT (24H ERROR IN TRANSFORMATION, 3X, I2, 3X, 11HZERO DIVIDE)
      GO TO 1330
  510 X(I) =  1.0/ XX(I)
      GO TO 740
C     TRANSFORMATION 8
  520 IF (ITRAN(8)) 530, 560, 530
  530 IT = 8
      ASSIGN 560 TO ILINE
      DO  550 I = 1, N
      ARG = D + XX(I)
      IF (ARG) 550, 540, 550
  540 PRINT 500, IT
      GO TO 1330
  550 X(I) = 1.0/ ARG
      GO TO 740
C     TRANSFORMATION 9
  560 IF (ITRAN(9)) 570, 610, 570
  570 IT = 9
      ASSIGN 610 TO ILINE
      DO  600 I = 1, N
      ARG = 1. - XX(I) ** 2
      IF (ARG) 580, 580, 600
  580 PRINT 590, IT
  590 FORMAT ( 24H ERROR IN TRANSFORMATION, 3X, I2, 3X, 47HZERO DIVIDE
     1 OR SQUARE ROOT OF A NEGATIVE NUMBER)
      GO TO 1330
  600 X(I) = ATAN (XX(I)/ SQRT (ARG))
      GO TO 740
C     TRANSFORMATION 10
  610 IF (ITRAN(10)) 620, 650, 620
  620 IT = 10
      ASSIGN 650 TO ILINE
      DO 640 I = 1, N
```

```
      ARG = 1. - XX(I)
      IF (ARG .GT. 0..AND. XX(I) .GE. 0.) GO TO 640
  630 PRINT 590, IT
      GO TO 1330
  640 X(I) = 2. * ATAN (SQRT (XX(I))/SQRT(ARG))
      GO TO 740
C     TRANSFORMATION 11
  650 IF (ITRAN(11)) 660, 690, 660
  660 IT = 11
      ASSIGN 690 TO ILINE
      DO 680 I = 1, N
      IF (E) 680, 670, 680
  670 PRINT  500, IT
      GO TO 1330
  680 X(I) = XX(I)/E
      GO TO 740
C     TRANSFORMATION 12
  690 IF (ITRAN(12)) 700, 720, 700
  700 DO 710 I = 1, N
  710 X(I) = SIN(XX(I))
      IT = 12
      ASSIGN 720 TO ILINE
      GO TO 740
C     TRANSFORMATION 13
  720 DO 730 I = 1, N
  730 X(I) = COS(XX(I))
       IT = 13
C     PRINT INPUT
  740 PRINT 750, JOB, FMT
  750 FORMAT (1H1, 10A8/ 23H FORMAT FOR SAMPLE DATA, 5X, 10A8)
      PRINT 760, NRUN, IRUN, (IFILE(I), I= 2, 19)
  760 FORMAT (15H NO. OF RUNS = , I2, 5X, 7HIRUN = ,  I2/
     121H TRANSFORMATION  1 = , I1, 5X, 20HTRANSFORMATION  2 = , I1, 5X,
     2 20HTRANSFORMATION  3 = , I1, 5X, 20HTRANSFORMATION  4 = , I1/
     321H TRANSFORMATION  5 = , I1, 5X, 20HTRANSFORMATION  6 = , I1, 5X,
     4 20HTRANSFORMATION  7 = , I1, 5X, 20HTRANSFORMATION  8 = , I1/
     521H TRANSFORMATION  9 = , I1, 5X, 20HTRANSFORMATION 10 = , I1, 5X,
     6 20HTRANSFORMATION 11 = , I1, 5X, 20HTRANSFORMATION 12 = , I1/
     721H TRANSFORMATION 13 = , I1, 5X,  9HMEDIUM = , I1/
     815H PLOT TYPE A = , I1,11X, 14HPLOT TYPE B = , I1,11X, 14HPLOT TYP
     9E C = , I1,11X, 14HPLOT TYPE D = , I1)
      PRINT 770, (IFILE(I), I = 20, 35), ITEST, IALPHA
  770 FORMAT ( 9H NOTIC = , I2, 16X, 6HIRA = , 13( I2, 1X)/ 7H NBR = ,
     1I2,18X, 5HNC = , I1, 20X, 8HITEST = , I2/24H ALPHA IDENTIFICATION
     2= , 13(5I1, 1X))
      IF (NC) 780, 800, 780
  780 PRINT 790, A, B, C, D, E
  790 FORMAT (29H TRANSFORMATION CONSTANT CARD, 3X, 4HA = , F14.6, 3X,
     1 4HB = , F14.6, 3X, 4HC = , F14.6/32X, 4HD = , F14.6, 3X, 4HE = ,
     2 F14.6)
  800 NM = N/2 + MOD (N,2)
      NN = 0
  810 NK = NN + 1
      NN = MIN0 (50, NM) +NN
C     PRINT ORIGINAL AND TRANSFORMED DATA
      PRINT 820,((XXP(J,I), XP(J,I), J = 1,2), I =NK,NN)
  820 FORMAT (1H1, 9X, 13HORIGINAL DATA, 15X, 16HTRANSFORMED DATA, 16X,
     113HORIGINAL DATA, 15X, 16HTRANSFORMED DATA/ (5X, F20.10, 10X, F20.
     210, 10X, F20.10,  10X, F20.10))
```

```
      NM = NM - 50
      IF (NM) 830, 830, 810
C     SORT DATA
  830 K = N - 1
      DO 850 I = 1, K
      II = I + 1
      DO 850 J = II, N
      IF (X(I) - X(J)) 850, 850, 840
  840 TEMP = X(J)
      X(J) = X(I)
      X(I) = TEMP
  850 CONTINUE
C     ESTABLISH PARAMETERS FOR PLOTS
      XMI = X(1)
      XMA = X(N)
      RANGE = XMA - XMI
      IF (NOTIC .EQ. 0) NOTIC = 15
      DX = RANGE / FLOAT(NOTIC)
      TEN (1) = 10.E-10
      IF (DX .GT. TEN(1)) GO TO 860
      DX = TEN(1)
      GO TO 900
  860 IF (DX .LT. 10.E10) GO TO 870
      DX = 10.E10
      GO TO 900
  870 DO 880 M = 2, 20
      TEN (M) = 10.** (M - 11)
      IF (DX .GT. TEN(M-1) .AND.DX .LE. TEN(M)) GO TO 890
  880 CONTINUE
  890 NDX = DX/ TEN(M-1) + .9
      DX = FLOAT (NDX) * TEN(M-1)
  900 DELTA = RANGE/ 500.
      FLN = FLOAT(N)
      SUM = 0.
      SUMSQ = 0.
C     COMPUTE AND PRINT MEAN ADD VARIANCE
      DO 910 I = 1, N
      SUM = SUM + X(I)
  910 SUMSQ = SUMSQ + X(I) * X(I)
      XBAR = SUM/ FLN
      SIGMA2 = (SUMSQ - FLN * XBAR ** 2)/ (FLN - 1.)
      SIGMA = SQRT (SIGMA2)
      PRINT 920, XMI, XMA, RANGE, N, XBAR, SIGMA2, SIGMA
  920 FORMAT (1H0, 8X, 10HMINIMUM = , E15.8, 20X, 10HMAXIMUM = , E15.8,
     110X,  8HRANGE = , E15.8/ 5X, 14HSAMPLE SIZE = , I5, 33X, 7HMEAN =
     2, E15.8/ 8X, 11HVARIANCE = , E15.8, 9X, 21HSTANDARD DEVIATION = ,
     3E15.8)
C     COMPUTE EMPIRICAL CDF
      DD = 0.
      J = 1
      IFREQ (J) = 0
      DO 990 I = 1, N
      IFREQ(J) = IFREQ(J) + 1
      IF (X(I) - X(I+1)) 930, 990, 930
  930 X(J) = X(I)
      U(J) = (X(J   - XBAR)/ SIGMA
      F(J) = FREQ(U(J))
      S(J) = FLOAT(I)/ FLOAT (N)
C     DETERMINE TEST STATISTIC D
```

```
      SN(J) = ABS (F(J) - S(J))
      IF (J - 1) 950, 940, 950
  940 SN1 (1) = ABS (F(J)
      GO TO 960
  950 SN1(J) = ABS (F(J)   S(J -1))
  960 ZMAX = AMAX1 (SN(J)  SN1(J))
      IF (ZMAX - DD) 980  )80, 970
  970 JJ = J
      DD = ZMAX
  980 J = J + 1
      IFREQ (J) = 0
  990 CONTINUE
C     PRINT STATISTICS
      J = J - 1
      II = J
      IK = 0
 1000 IL = IK + 1
      IK = MINO (50, II) + IK
      PRINT 1010
 1010 FORMAT (1H1, 54X, 29HKOLMOGOROV TEST FOR NORMALITY//3X,      11HOR
     1DER INDEX, 6X, 11HTRANSFORMED,18X, 12HSTANDARDIZED, 1X,
     2 15HTHEORETICAL CDF, 2X, 12HOBSERVED CDF, 3X,          14HABSOLUT
     3E VALUE, 3X, 14HABSOLUTE VALUE/ 7X, 3H(J), 11X, 9HDATA X(J),
     4 7X, 9HFREQUENCY, 4X,
     512HVARIATE U(J), 5X, 4HF(J), 12X, 4HS(J), 9X, 11HF(J)-S(J-1), 7X,
     69HF(J)-S(J)/)
      PRINT 1020,((I,   X(I), IFREQ (I), U(I), F(I), S(I), SN1(I), SN(I)
     1), I =IL,IK)
 1020 FORMAT        (6X, I4, 6X, F14.6, 10X, I4, 4X, F14.11, 1X, F14.12,
     1 2X, F14.12, 1X, F14.13,  2X, F14.13)
      II = II - 50
      IF (II) 1030, 1030, 1000
 1030 PRINT 1040, JJ, DD
 1040 FORMAT ( 3HOD(, I4, 4H) = , F7.4)
      T(1) = XMI
C     COMPUTE THEORETICAL CDF
      DO 1050 L = 1, 500
      UV = (T(L) - XBAR)/SIGMA
      FN(L) = FREQ(UV)
      IF (T(L) .GT. XMA) GO TO 1060
      T(L+1) = T(L) + DELTA
 1050 CONTINUE
 1060 DO 1320 II = 1, 5
C     OBTAIN VALUE FOR D SUB ALPHA OF N
      IF (IALPHA (II, IT)) 1070, 1320, 1070
 1070 IF (IA .EQ. 0 .AND. IB .EQ. 0 .AND. IC .EQ. 0) GO TO 1150
      IF (N - 30) 1080, 1080, 1090
 1080 NN = N - 3
      DALPHA = (TABLE (NN, II))
      GO TO 1100
 1090 DALPHA = TABLE (28, II)/ SQRT(N)
 1100 IF (IB .EQ. 0 .AND. IC .EQ. 0) GO TO 1120
C     COMPUTE CONFIDENCE LIMITS
      DO 1110 I = 1, J
      SU(I) = S(I) + DALPHA
 1110 SL(I) = S(I) - DALPHA
 1120 IF (IA) 1130, 1150, 1130
C     COMPUTE REJECTION REGION
 1130 DO 1140 I = 1, 500
```

```
      FU(I) =FN(I) + DALPHA
 1140 FL(I) =FN(I) - DALPHA
C****P L O T S
C     HEADINGS
 1150 DO 1310 KK = 1,4
      IF (IFILE (KK + 15)) 1160, 1310, 1160
 1160 PRINT 1170, JOB, IRUN, N
 1170 FORMAT (2H$2, 15X, 10A8/ 1H$, 15X, 11HRUN NUMBER , I2, 6H WITH ,
     1I5, 13H OBSERVATIONS)
      GO TO (1180, 1200, 1220, 1240), KK
C     HEADINGS FOR PLOT TYPE A
 1180 PRINT 1190, ALPHA(II), DALPHA
 1190 FORMAT (1H$,15X, 39HOBSERVED RELATIVE CUMULATIVE FREQUENCY,/ 1H$,
     115X, 46HTHEORETICAL RELATIVE CUMULATIVE FREQUENCY, AND/ 1H$,
     215X, 34HCRITICAL REGION BOUNDS FOR ALPHA =, F5.2, 3X,
     3 22HLEVEL OF SIGNIFICANCE,/1H$, 15X, 11HD(ALPHA) = , F5.3)
      GO TO 1260
C     HEADINGS FOR PLOT TYPE B
 1200 ICB = 100. * (1. - ALPHA(II)) + .5
      PRINT 1210, ICB, DALPHA
 1210 FORMAT (1H$,15X, 39HOBSERVED RELATIVE CUMULATIVE FREQUENCY,/ 1H$,
     115X, 42HTHEORETICAL RELATIVE CUMULATIVE FREQUENCY,/      1H$,
     215X, I2,1X,78HPERCENT CONFIDENCE BANDS FOR THE THEORETICAL CUMULAT
     3IVE DISTRIBUTION FUNCTION,/ 1H$,15X, 11HD(ALPHA) = , F5.3)
      GO TO 1260
C     HEADINGS FOR PLOT TYPE C
 1220 ICB = 100. * (1. - ALPHA(II)) + .5
      PRINT 1230, ICB
 1230 FORMAT (1H$,15X, 46HOBSERVED RELATIVE CUMULATIVE FREQUENCY AND THE
     1/ 1H$,15X,I2, 1X,77HPERCENT CONFIDENCE BANDS FOR THE THEORETICAL C
     2UMULATIVE DISTRIBUTION FUNCTION)
      GO TO 1260
C     HEADINGS FOR PLOT TYPE D
 1240 PRINT 1250
 1250 FORMAT ( 1H$,15X, 54HOBSERVED AND THEORETICAL RELATIVE CUMULATIVE
     1FREQUENCY)
C     PLOT MEAN AND S(X)
 1260 CALL GRF (XMI, 0., XMA, 1., XMI, 0.,DX,.05,IRA(IT),2,0, 1, XBAR,
     1 .5, 1, 1, 2H12, -1, 0)
      CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1, J, X(1),
     1 S(1), 1, 1, 2H75, 0, 0)
      GO TO (1270, 1280, 1290, 1300), KK
C     ADDITIONAL CURVES FOR PLOT TYPE A
 1270 CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1,1000,T(1),
     1FN(1), 1, 1, 2H75,-1, 0)
      CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1,1000,T(1),
     1FL(1), 1, 1, 2H75,-1, 0)
      CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1,1000,T(1),
     1FU(1), 1, 1, 2H75,-1, 0)
      GO TO 1310
C     ADDITIONAL CURVES FOR PLOT TYPE B
 1280 CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1,1000,T(1),
     1FN(1), 1, 1, 2H75,-1, 0)
      CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1, J, X(1),
     1SL(1), 1, 1, 2H75, 0, 0)
      CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1, J, X(1),
     1SU(1), 1, 1, 2H75, 0, 0)
      GO TO 1310
C     ADDITIONAL CURVES FOR PLOT TYPE C
```

```
 1290 CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1, J, X(1),
     1SL(1), 1, 1, 2H75, 0, 0)
      CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1, J, X(1),
     1SU(1), 1, 1, 2H75, 0, 0)
      GO TO 1310
C     ADDITIONAL CURVES FOR PLOT TYPE D
 1300 CALL GRF (XMI, 0., XMA, 1., XMI, 0., DX,.05,IRA(IT),2,1,1000,T(1),
     1FN(1), 1, 1, 2H75,-1, 0)
 1310 CONTINUE
 1320 CONTINUE
 1330 IRUN = IRUN + 1
      IF (IRUN - NRUN) 1340, 1340, 1350
 1340 GO TO ILINE
 1350 CALL INTVL (Z)
 1360 FORMAT (19H0TIME IN SECONDS = , F10.2)
      PRINT 1360, Z
      GO TO 100
      END
T         SUBTYPE,DATA
```

VIII. REFERENCES

1. Birnbaum, Z. W. (1952), Numerical Tabulation of the Distribution of Kolmogorov's Statistic for Finite Sample Size, Journal of the American Statistical Association, Vol. 47, pp. 425-441.

2. Cochran, W. G. (1954), Some Methods for Strengthening the Common χ^2 Tests, Biometrics, Vol. 10, pp. 417-451.

3. Kac, M.; Kiefer, J.; and Wolfowitz, J. (1955), On Tests of Normality and Other Tests of Goodness of Fit Based on Distance Methods, The Annals of Mathematical Statistics, Vol. 26, pp. 189-211.

4. Lilliefors, H. W. (1967), On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown, Journal of the American Statistical Association, Vol. 62, pp. 399-402.

5. Massey, F. J. (1951), The Kolmogorov-Smirnov Test for Goodness of Fit, Journal of the American Statistical Association, Vol. 46, pp. 68-78.

6. Miller, L. H. (1956), Table of Percentage Points of Kolmogorov Statistics, Journal of the American Statistical Association, Vol. 51, pp. 111-121.

7. Slakter, M. J. (1967), A Comparison of the Pearson Chi-Square and Kolmogorov Goodness-of-Fit Tests with Respect to Average Power, Personal Communication, State University of New York at Buffalo.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* U. S. Naval Weapons Laboratory	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

A COMPUTER PROGRAM FOR THE KOLMOGOROV GOODNESS OF FIT TEST FOR NORMALITY

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*

Carl B. Bates and Jacqueline R. Orsulak

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
January 1968	49	

8a. CONTRACT OR GRANT NO. b. PROJECT NO.	9a. ORIGINATOR'S REPORT NUMBER(S) TM K-2/68
c. d.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

Following a brief discussion on the background, applicability, and limitations of the Kolmogorov Test, a computer program of the Kolmogorov Test for normality is described. The program is applicable for testing the null hypothesis that a random sample is from a parent normal population whose mean and variance are equal to those of the sample distribution. The program determines the minimum and maximum sample values, computes sample estimates of the mean and variance of the hypothesized normal distribution and computes the Kolmogorov statistic. Five optional significance levels (0.20, 0.15, 0.10, 0.05 and 0.01) are available, and sample size limitations are $4 \leq n \leq 2500$. An optional feature provides for CRT plot output applicable for both hypothesis testing and interval estimation.

The program is coded in FORTRAN IV for the IBM 7030 (STRETCH) computer.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408